DOCUMENT RESUME

ED 286 831 SP 029 303

AUTHOR McCarthy, Jean
TITLE Construction of Tests in the Cognitive and Psychomotor Domains for Skin and Scuba Diving.
PUB DATE 87
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cutting Scores; Higher Education; Knowledge Level; *Mastery Tests; Psychomotor Skills; *Test Reliability; *Test Validity; *Underwater Diving

ABSTRACT

The fundamental purposes of this study were to develop mastery tests in the cognitive and psychomotor domains for skin and scuba diving and to establish validity and reliability for the tests. A table of specifications was developed for each domain, and a pilot study refined the initial test batteries into their final form. In the main study, two-trial cognitive data were collected from 104 male and female college students enrolled in a skin and scuba diving course. Data were analyzed for test reliability and validity. Two skin diving and three scuba diving skills tests were administered to 98 subjects in two trials. One-way analysis of variance intraclass correlations were used to ascertain test-item reliability and validity. Both cognitive and psychomotor tests were found to be valid and reliable. Cutting scores were set for each test. Results showed the tests to be administratively feasible. (Author/MT)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED286831

Construction of Tests in the Cognitive and Psychomotor Domains for Skin and Scuba Diving

Jean McCarthy

Mankato State University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. McCarthy

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

SP029303

CONSTRUCTION OF TESTS IN THE COGNITIVE AND PSYCHOMOTOR DOMAINS FOR SKIN AND SCUBA DIVING.

The purpose of this study was to construct mastery tests in the cognitive and psychomotor domains for sport diving. In addition, the study was designed to determine validity and reliability and to evolve workable and practical scoring and administrative procedures for the tests. Tables of specification for each domain, cognitive and psychomotor, and subsequent pilot study of initial test batteries refined the tests into their final forms for analysis. Two-trial data were collected from 104 male and female students enrolled in a college-level skin and scuba diving course and analyzed for test reliabilities and validities. Reliability claims for the alternate forms of the cognitive test were supported by a coefficient of equivalence (r=.784) and Cronbach Alpha coefficients of homogeneity (r=.772, r=.792). Content validity derived from item generation from an authority-based table of specifications and from an item analysis that showed the majority of the items to be within limits recommended by experts for a mastery test. Reliability for the two skin-diving tests (underwater distance swim, mask-and-snorkel clearance) and the three scuba-diving motor tests ("bail-out," "ditch-and-don," "buddy switch") was determined by intraclass correlations (r=.797 to r=.815). Content validity for the five motor tests was established by an authority-based table of specifications, and construct validity by nonsignificant *t* tests of means between the main study's subjects and a sample of experienced divers. Further, low inter-item relationships (r=-.138 to r=.571) were observed. Cutting scores for domain mastery were suggested. Results produced tests in the cognitive and psychomotor domains for sport diving that can evaluate diving skills adequately with one trial and that are an administratively feasible mastery test battery.

Construction of Tests in the Cognitive and Psychomotor Domains for Skin and Scuba Diving

Since 1943 when Cousteau and Gagnan developed the first modern scuba lung, diving of all kinds has enjoyed meteoric popularity in the United States. Today there are approximately nine million active sport divers in this country. About 80 percent of these divers have received some kind of formal training from any of the several national diving instruction agencies, from schools and dive shops scattered across the country which offer courses, or from on-site lessons at diving-oriented resorts. Virtually all of these sport diving courses have little or no standard evaluation procedures for either the cognitive or psychomotor domains. This dearth of standardized testing has resulted in a wide variety and quality of provincial evaluation techniques. Since a majority of the people teaching sport diving in the United States has not received professional preparation in testing educational outcomes, the validity and reliability of the myriad of teacher-constructed tests are suspect. Very few professional educators are teaching sport-diving courses. Many schools and colleges hire independent diving instructors or instructors from dive shops as adjunct faculty to provide courses for their students. With evaluation of diving performance in question in many courses, concern should be focused on the safety of those students certified from such courses and on the integrity of the sport as an entity in the mosaic of recreational activities available to the public.

Need for the Study

Examination of the course manuals (NAUI, 1984; PADI, 1980; YMCA, 1978; Zimos and Hammes, 1984) developed by the diving instruction

agencies revealed only general directions and suggestions for testing diving skills and knowledge. A 1984 computer search of the data banks of ERIC, SPORT and Dissertation Abstracts along with a review of the microfiche file for studies catalogued by the University of Oregon and the index of Completed Research in Health, Physical Education and Recreation produced no standardized tests in either the cognitive or psychomotor domains of sport diving. Tests in the two domains should be developed in the interest of (1) identifying mastery states for certification purposes, (2) providing standardized performance criteria for comparison of instructional and research results, and (3) supplying evaluative guidelines in test construction and use in sport diving.

Obviously a sizeable amount of research must be done to fill this test void in diving. The logical source of this research is the academic community where people are trained in the discipline of systematic study, are equipped with the hardware and software to conduct the research, and are inclined by virtue of their tradition and professional commitment to examine problems and attempt solutions. Somers (1979) endorsed this need when he stated:

> The next phase in the growth and development of the scuba diving industry must be cultivated in the academic community. The industry is badly in need of the direction, prestige, and research that can and should be provided by professional educators and scientists. The academic community must step forth to meet this challenge and the scuba diving community must accept and encourage this involvement (p. 50).

Procedures

The fundamental purposes of this study were to develop mastery tests in the cognitive and psychomotor domains for skin and scuba sport diving and to establish validity and reliability for all tests.

Also, it was intended to evolve workable and practical administrative and scoring procedures for them.

Based upon information supplied by several authoritative sources such as the training directors of the national diving instruction agencies, the sport diving course manuals developed by those agencies, standard texts and reference books in sport diving (CNCA, 1985; Graver, 1978; Hammes and Zimos, 1980; Ketels and McDowell, 1975; NOAA, 1975; Navy Department, 1984), the American National Standards Institute (1972), and active diving instructors, a table of specifications was made for each domain, cognitive and psychomotor. From these two tables, whose final contents were verified by the agency training directors, derived the initial forms of the cognitive and motor tests. A pilot study with 49 subjects similar to those used in the main study resulted in modifications to both the cognitive and motor test batteries. The refined test batteries were then administered on a test-retest basis to 104 volunteer subjects enrolled in a college course in skin and scuba diving. The motor test battery was also given to 10 experienced divers. The resulting data were analyzed to provide information for determining test-item reliabilities and validities for both batteries and cutting scores on all tests to indicate domain mastery. All subjects read and signed an informed consent form prior to participation in the testing.

Cognitive Tests

Using the table of specifications as a guide, a pool of 300, four-foil, multiple-choice questions was written. Each of 150 questions was composed item for item with a similar but not identical counter-

part question designed to test the same information bit on two test forms, A and B. Items were written to test three levels of cognitive behavior: knowledge of facts, comprehension of concepts, and application of principles. Item analysis of the test data for difficulty, discrimination, and distracter function in the pilot study eliminated the 50 weakest item pairs from the two test forms. These eliminated items reduced the original item pool to the 200 statistically best items of 100 matched pairs on Form A and Form B. Inter-form equivalence prior to removal of the 50 pairs of items was $r = .865$; the equivalence coefficient after removal of the 50 item pairs was $r = .884$. This demonstrated improved inter-form reliability with a shorter test but better items.

The main study subjects took the two forms of the cognitive-domain test in sport diving on two successive class days at the end of the quarter of their enrollment in the diving course. The order of test-form taking was determined by a toss of a coin for each class. Test data were scanned and then analyzed by the Mankato State University Computer Services' optical scanner and Sperry Univac 1100/80 computer with SPSS programs.

The descriptive statistics of the subjects' performances on Form A and Form B taken in random order are presented in Table 1.

Table 1

Cognitive Test Performance (N=104)

Statistic	Form A	Form B
Range	52.93	51.95
Mean	74.71	74.70
Median	75.50	76.00
Mode	78.00	74.00
Standard Deviation	8.34	8.41
Standard Error	.82	.82
Skewness	-.41	-.37
Kurtosis	.21	-.43

The data showed remarkably similar results, a desirable characteristic for alternate forms of the same test. Equivalent forms reduce the problem of retesting to mastery. Form equivalence was further supported by a strong correlation of r=.784 which became r=.879 when predicted by the Spearman-Brown prophecy formula for the 200 items on both forms.

Intra-form consistency was also examined by computing the Cronbach Alpha coefficient. This indicator of internal homogeneity was .772 for Form A, .792 for Form B, and .878 for the combined forms.

The consistency of responses across parallel test forms was studied with the McNemar test of change to determine the reliability with which respondents answered matched items. Fifty of the 100 item pairs produced non-significant chi-square values, suggesting consistent

understanding or misunderstanding of the same item bit. The other 50 item pairs' inconsistent responses across test forms can be explained in part by a variety of interfering variables such as guessing, learning between testings, differing item sophistication and an imperfect equivalence correlation.

Item difficulty indices on a mastery test are useful because they indicate what material from the domain has been learned. The mean difficulty index for both forms was .747 with the range of indices on Form A .010 to .779 (difficult to easy) and .019 to .760 on Form B. Slightly higher indices than the optimum .500 are recommended on a mastery test to reinforce important information (Brown, 1976). The modal range on Form A was .600 to .700, and the modal range on Form B was .700 to .800.

Item discrimination indices are an internal test criterion helpful in mastery testing since only rarely are there external criteria available for item comparison. High, positive indices above .15 indicate questions answered correctly more often by respondents from the top-scoring group than by the low-scoring group. The mean discrimination index for Form A was .199 and for Form B was .208. Both forms produced seven questions each with negative indices which can be attributed either to poor discrimination qualities or to being too easy or too difficult, a not wholly unexpected situation for a mastery test where respondents are expected to do well.

More precise information about the discriminating power of the test can be obtained from biserial correlations between each item score and the test score. Safrit (1981) has indicated that coefficients

of +.30 or greater are desired. Form A had an average item biserial correlation of +.349 with a modal range of .300 to .400. Form B had an average coefficient of +.398 with a modal range of .500 to .600. The distribution of biserial coefficients paralleled the discrimination indices on the two forms as anticipated on a mastery test.

The item analysis also identified the distracters of the multiple-choice questions that were not working. Seventy-six questions on Form A and 65 questions on Form B had all distracters working. Twenty questions on Form A and 23 questions on Form B had one distracter that was not functioning. The remaining four questions on Form A and 12 questions on Form B had two distracters not chosen. All of the questions with two distracters unchosen had difficulty indices of .900 or larger, indicating easy questions appropriate for mastery tests where students are expected to do well on a test with deliberately skewed difficulty.

Validity

Validity claims for the alternate forms of the cognitive test are derived from the generation of a pool of questions developed from a table of specifications approved by diving authorities and from an item analysis that showed the majority of questions on each form to be within limits recommended by experts. The test appears to have content validity and possesses good precision in the measurement of cognitive skills in sport diving.

Reliability

Reliability claims for the alternate forms of the cognitive test were supported by strong coefficients of equivalence, high Cronbach

Alpha coefficients of homogeneity and a fair degree of inter-form matched-item consistency reflected by McNemar values. The two forms of the test appear to be quite similar and measure cognitive skills in sport diving reliably.

Motor Tests

A list of motor skills perceived as essential to safe sport diving was developed from a table of specifications and verified by diving authorities. The skills were integrated to form four tasks that would lend themselves to measurement, two in skin diving and two in scuba diving. These tasks are widely used by diving instructors as both instructional drills to teach the skills involved in them and as tests of those skills. Subject-selected equipment was used with a weight belt optional in the scuba items. Standard directions were developed for each test and were employed by the testers at all testing sessions. Two trials of all tests were administered on each of two successive class meeting days at the end of the academic quarter of enrollment in the diving course.

Skin Diving Tests

One skin-diving skill test was an underwater distance swim with a push-off start in a 25-yard pool wearing conventional skin diving equipment. The trial score was the total distance covered rounded to the nearest yard. The other skin-diving skill test was a timed, one-breath, mask and snorkel clearance from the bottom of a 12-feet deep pool while wearing foot fins. The trial score was in seconds rounded to the nearest tenth.

Scuba Diving Tests

One of the scuba diving skills tests was the task known as "bail-out." It is a timed event involving the diver making a foot-first entry in the deep end of a pool carrying all diving equipment rather than wearing it. The equipment is then put on properly prior to surfacing. The trial score was the elapsed time to the nearest second.

The other initial scuba-diving skills test was the task known as "ditch and don." A fully equipped diver surface dived to the bottom in the deep end of a pool; took off all equipment, leaving it on the bottom; and made a free ascent to the surface. After no more than a two-minute surface interval of treading water, the diver descended to the equipment and donned it all properly. At that point he signaled his buddy diver to surface dive down to him and buddy-breathe from the scuba lung with him. The trial score was the sum of the elapsed times to the nearest second from each of the two underwater phases of the task.

Pilot Study

Pilot study of the four motor tasks revealed several minor scoring problems which were remedied prior to the main study. It was also decided to separate the buddy breathing from the "ditch and don" task and add it in a task known as the "buddy-switch" to avoid confounding effects. This task entails a completely equipped scuba diver and a buddy-diver with no equipment buddy-breathing from the shallow end of a pool to the deep end where they exchange all of the diving equipment and return to the shallow end of the pool before surfacing. The trial score was the elapsed time to the nearest second from submersion to surfacing.

More pilot study was made of the five revised diving skills tests on two successive class meeting days at the end of the next academic quarter. The refined protocol worked well and the integrity of the skills derived from the table of specifications was preserved. Correlation analysis of the test-retest data and simultaneous testing data for reliability and objectivity purposes suggested no further test refinements were needed. See Table 2 for the pilot study findings.

Table 2

Pilot Study Reliability and Objectivity Coefficients (N=47)

Test Item	Reliability	Objectivity
Underwater Distance Swim	.779	.999
Mask and Snorkel Clearance	.751	.995
Bail Out	.685	.999
Ditch and Don	.702	.999
Buddy Switch	.632	.999
Bail Out + Ditch & Don + Buddy Switch	.861	.999

Main Study

Test-retest trials of the five sport-diving skills tests were administered to 98 subjects who provided complete and usable data for computer analysis. In general, the descriptive statistics of the data from two trials of each test item were similar but demonstrated some variation as the tasks became more complex.

One-way ANOVA intraclass correlations were computed on the trials

to ascertain test-item reliability. Table 3 has the item coefficients.

Table 3

Diving Skills Test-Retest Intraclass Correlations (N=98)

Test Item	R
Underwater Distance	.815
Mask and Snorkel	.797
Bail Out	.799
Ditch and Don	.811
Buddy Switch	.810

Intercorrelations among test-item scores and composite test scores reveal information about predictive accuracy and item weight. Weak correlations among the test items suggest that each item measures a unique aspect of the total skill used as the performance standard.

The intercorrelations between the two skills tests in skin diving, taken separately or collectively, were not different from zero. The coefficients of determination showed the test items shared 1.9 percent or less of their variance in common. Due to different score units for the two skin diving skills tests, a composite score was not computed.

As in the skin-diving battery, the intercorrelations among the scuba tests were weak. They ranged between r=.136 to r=.571 for both trials and the sum of the tests and trials. By contrast, the item-to-total score associations were strong. They ranged between r=.615 to r=.843 for both trials and the sum of the tests and trials. Except for one coefficient, the item-to-total correlations exceeded the .70 suggested

by Mathews (1978) as necessary to indicate prediction precision.

Stepwise, multiple regression analysis was made of the separate and summed trials of the scuba skills tests to identify the variance contributions by the test items to the prediction of the battery composite score. Tables 4, 5 and 6 have the regression analysis information.

Table 4

Stepwise Regression and Multiple Correlation of Scuba Test First Trials With Total Score (N=98)

Test	Multiple R	R^2
Buddy Switch	.801	.642
Ditch and Don	.969	.938

$\hat{y}$ (Sum) = 7.919 + 1.117 BS_1 + 1.362 DD_1

Table 5

Stepwise Regresesion and Multiple Correlation of Scuba Test Second Trials With Total Score (N=98)

Test	Multiple R	R^2
Ditch and Don	.816	.666
Buddy Switch	.946	.895

$\hat{y}$ (Sum) = 36.981 + 1.354 DD_2 + .969 BS_2

Table 6

Stepwise Regression and Multiple Correlation of Summed Scuba Test Trials With Summed Total Scores (N=98)

Test	Multiple R	R^2
$DD_1 + DD_2$	.843	.710
$BS_1 + BS_2$	.967	.934

$\hat{y}\ (Sum_1 + Sum_2) = 36.683 + 1.397\ (DD_1 + DD_2) + 1.039\ (BS_1 + BS_2)$

DD = Ditch and Don

BS = Buddy Switch

Examination of the regression analysis tables reveals consistent, strong associations between the "ditch-and-don" and "buddy-switch" items with the total score, indicating enhanced prediction properties and item weight analysis. The "bail-out" test persistently explained less than 10^{-30} of the variance. Second trials or summed trials added little information to what is produced from first trials of the "buddy-switch" and "ditch-and-don" scuba test items.

The final analysis of the motor test-item data in sport diving was a test of mean differences between the averaged two-trial performance scores of the main study subjects and the single trial performance scores of 10 experienced divers who volunteered to be tested under the same conditions. Independent t tests were computed for the two groups on the five tasks and none of the mean comparisons resulted in significant differences ($p > .05$). This finding suggests that experienced divers performed the tasks similarly to the novice divers in the study and that experienced divers possess diving skills that persist over time from their original instruction as a result of ongoing diving experiences.

Validity

Claims of content validity derive from the five motor tasks being identified from an authority-based table of specifications. Logical validity arises from confirmation by diving authorities that these tasks are what divers do. Construct validity was demonstrated by the non-significant differences between the study's subjects and a group of experienced divers. Inter-correlation and regression analysis showed item weight and component predictor precision to be good.

Reliability

The motor tests' reliability was confirmed by strong intraclass correlations ($R > .800$). A stepwise regression analysis of trials and summed trials showed a consistent pattern of item contributions to the test battery total score.

Cutting Scores

Both the cognitive test forms and the motor skills tests were designed for use as mastery tests in sport diving. Cutting scores that divide the acceptable and unacceptable performances were set at the means of each test. This was done after considering that the study showed no differences on the motor tests between the novice and experienced divers; the tests' means reflected reasonable performance levels for safety and enjoyment; cutting scores set in the middle ranges of the respondents' scores will maximize the number of correct selection decisions; and the mean is a representative and stable statistic for repeated performances. Table 7 has the cutting scores rounded to the nearest useful place. Subsequent experience with these cutting scores produced an overall

first-trial success of 53.9 percent and an eventual success ratio after repeated trials of 82.7 percent.

Table 7

Cutting Scores on Sport Diving Cognitive and Motor Tests

Test	Cutting Score
Written Test (A or B)	75 pts.
Underwater Distance Swim	52 yds.
Mask and Snorkel Clearance	12.7 secs.
Bail Out	84 secs.
Ditch and Don	150 secs.
Buddy Switch	170 secs.

Conclusions

Based on the findings of this study and bound by the constraints of its design, the following conclusions were made.

1. The goal of developing tests in the cognitive and psychomotor domains for sport diving with acceptable validities and reliabilities was achieved.

2. The protocol of this study provides reasonable and workable guidelines for the construction of cognitive and motor tests in diving.

3. The alternate cognitive test forms are sufficiently equivalent to be used interchangeably.

4. The "bail-out" scuba skills test does not contribute significantly to the assessment of diving motor behavior.

5. Single trials of the skills test items are sufficient for evaluation in the psychomotor domain of diving.

6. The cognitive and motor test batteries in sport diving developed in this study appear to be administratively feasible in a typical school setting and scientifically defensible.

REFERENCES

American National Standards Institute (1972). Minimum course content for safe scuba diving instruction. New York: American National Standards Institute.

Brown, G. (1976). Principles of educational and psychological testing. New York: Holt, Rinehart, Winston.

Council for National Cooperation in Aquatics (CNCA) (1985). The new science of skin and scuba diving. Piscataway, N.J.: Association Press.

Graver, D. (1978). Professional association of diving instructors dive manual. Santa Ana, Ca.: Professional Association of Diving Instructors.

Hammes, R. and Zimos, A. (1980). Safe scuba. Long Beach, Ca.: National Association of Scuba Diving Schools.

Ketels, H. and McDowell, J. (1975). Safe skin and scuba diving. Boston: Little, Brown and Company.

Mathews, D. (1978). Measurement in physical education. Philadelphia: W. B. Saunders.

National Association of Underwater Instructors (NAUI) (1984). Professional resource organizer manual. Montclair, Ca.: National Association of Underwater Instructors.

National Oceanic and Atmospheric Administration (NOAA) (1975). NOAA diving manual. Washington, D.C.: National Oceanic and Atmospheric Administration.

Navy Department (1984). U.S. navy diving manual. Washington, D.C.: Navy Department.

Professional Association of Diving Instructors (PADI) (1980). Standards and procedures manual. Santa Ana, Ca.: Professional Association of Diving Instructors.

Safrit, M. (1981). Evaluation in physical education. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Somers, L. (1979). Underwater education. Journal of Physical Education and Recreation, May, 50.

Young Men's Christian Association (YMCA) (1978). YMCA underwater activities leadership manual. Norcross, Ga.: National YMCA Underwater Activities Program.

Zimos, A. and Hammes, R. (1984). Gold book. Long Beach, Ca.: National Association of Scuba Diving Schools.